多位态研究主题识别及其演化路径方法研究*

王康[1] 高继平[2] 潘云涛[2] 陈悦[1]
[1] 大连理工大学科学学与科技管理研究所暨 WISE 实验室 大连 116024 [2]中国科学技术信息研究所 北京 100038

摘 要：[目的/意义]科学主题演化路径对认识科学发展过程和预测未来发展趋势具有重要的意义。鉴于现有研究将演化路径上的主题同等看待的缺陷，提出一种新的多位态科学主题识别及其演化路径方法。[方法/过程]以向心度和密度将每个时间区间的主题划分为核心－成熟型、边缘－成熟型、边缘－非成熟型和核心－非成熟型 4 种类型，并利用余弦相似度将不同时间区间的主题进行关联，以展示不同类型科学主题之间的动态交叉演化关系。[结果/结论]以 3D 打印领域文献为例，从技术发展阶段、主题识别、主题类型划分和主题演化路径 4 个方面对 3D 打印技术发展过程进行测度，结果证明该方法对科学主题识别及其演化路径展示具有较好效果。

关键词：主题识别 主题演化 战略坐标 主题相似度 3D 打印
分类号：G306.1
DOI：10.13266/j.issn.0252－3116.2021.11.012

1 引言

随着科学技术的快速发展，文献增长速度日益加快，面对海量的科研成果如何快速有效地发现研究热点和前沿是图书情报、科学学等领域关注的重要议题。普赖斯认为：为了更好地理解当今社会的科学领域，必须沿着科学历史发展的轨迹去追溯，抓住关键的转折。科技文献不仅承载着科学技术的发展历史，同时还承接指引着人类继续认识世界和改造世界。数据密集型科研范式[1]下的科技文献主题演化路径挖掘不仅能够帮助科研人员快速掌握领域发展概况，还能为其预测未来发展方向提供重要的依据。目前，以文献关键词为单元的主题演化路径分析主要从“时间－主题聚类”和在此基础上更深一步的“时间－主题聚类－主题相似度”两个角度展开。“时间－主题聚类”即将目标文献集按照时间顺序划分为各个区间，然后对每个区间文献的关键词进行提取并聚类分析，如敦帅等[2]对中国企业可持续发展研究的文献进行关键词共现的时序分析以揭示其态势演进与主题演化；杜丽君等[3]以共现网络图的时序变化分析情报学和计算机科学的交叉主题演化规律；王璐瑶等[4]对《科学学与科学技术管理》创刊 40 年来发表的研究性论文进行分段关键词聚类、突现和战略坐标分析，全景展现了与我国科技发展阶段相对应的发展的时代特征、总体规律与突出贡献；Y. Wu 等[5]从 Web of Science 数据库中选择 10 个影响因子最高的精神病学期刊的所有文章，将其分为 3 个时期并使用信息可视化软件 Sci2 进行共词和聚类分析，用于精神病学研究重点和发展趋势的初步探索；O. B. Onyancha[6]对 1971 年至 2015 年每 10 年图书馆和信息科学（LIS）领域发表研究文章的关键词进行聚类分析以确定 LIS 中最突出、最常见的研究主题以及主题的演变方式；X. Y. Han[7]将 1996－2019 年 LIS 期刊文章划分为 5 个阶段，利用 LDA 模型从 14 035 篇文档中识别基本主题，研究 LIS 的发展；C. Huang 等[8]等构建教育领域核心论文的关键词共现网络，分析各个阶段网络密度的变化以确定教育领域核心主题的发展趋势。

“时间－主题聚类－主题相似度”即在“时间－主题聚类”的基础上将相邻时间区间的主题进行相似度分析，以此建立不同时区的主题关联，用以表征主题演化关系。如牛力等[9]根据关键词共现关系，利用社区探测算法对不同时间段关键词进行主题聚类，并通过社区之间节点的重合情况勾勒中国档案学研究内容的动态变化；王晓光等[10]根据社区主题表示算法和社区相似度匹配算法，构建了一个科研主题演化分析模型；

* 本文系国家重点研发计划项目“颠覆性技术识别理论、方法与专家预判系统”（项目编号：2019YFA0707201）研究成果之一。
作者简介：王康（ORCID：0000－0001－9170－6065），博士研究生；高继平（ORCID：0000－0003－4965－7065），副研究员，博士；潘云涛（ORCID：0000－0002－9930－9416），研究员，硕士；陈悦（ORCID：0000－0001－5272－9459），教授，博士，博士生导师，通讯作者，E-mail：chenyuedlut@163.com。
收稿日期：2020－12－07 修回日期：2021－02－24 本文起止页码：113－122 本文责任编辑：徐健

谭春辉等[11]利用LDA主题模型进行主题识别,然后利用余弦相似度对数据挖掘领域进行主题路径演化分析;颜端武[12]运用LDA主题模型对相邻时间窗口主题之间的相似度进行时序主题关联演化分析;周耀林等[13]利用社区探测算法进行主题聚类,并在此基础上利用数字化基础设施平台CorTexT[14]分析国际图情领域研究主题随时间演化态势;赵跃等[15]利用同样的方法研究了改革开放40年来中国档案学的核心主题演化轨迹。国外较多学者同样利用上述思路进行各学科、领域的主题演化研究,如A. Rule等[16]提出了一种自动文本分析策略,其对1790－2014年总统国情咨文语料库进行主题聚类,并将这些主题编织在一起,构建河流网络以捕获美国历史上的政治话语流,揭示美国政治演变的概况;F. L. Matos等[17]在利用Louvain算法主题聚类后,通过创建跨时间网络以桑基图的形式绘制了海底峡谷研究的地图,确定峡谷研究中现有的知识群、历史趋势和新兴话题;A. Marvuglia等[18]利用同样的方式动态评估了城市可持续方面的进展与挑战。

上述方法在刻画主题演化路径方面具有重要的方法论意义,但其忽略了对各主题相对重要性的分析,即将主题同等对待在进行科技政策制定过程中尚有缺憾之处。事实上,依据在主题网络中的位置及其发展状态,各主题是有位态区分的,即有核心、边缘、成熟、新兴等类型之分,明确不同类型主题及其之间的动态交互转化关系对认识领域发展过程和未来趋势具有重要的作用。岳丽欣等[19]利用LDA主题模型将我国医疗健康信息领域主题分为核心主题和边缘主题,并在此基础上进行主题相似度演化分析,但该方法仅依据主题聚类团体的大小划分核心－边缘主题,无法更细粒度地对每个主题所处的位态进行区分。基于此,笔者在主题路径研究中引入战略坐标法,通过向心度和密度对不同时间段主题类型进行划分,继而利用余弦相似度关联相邻时间段的主题,据此提出一种新的多位态科学主题识别及其演化路径方法,用以揭示核心、边缘、成熟、新兴等类型主题之间的动态演化关系,形成"时间－主题聚类－主题位态－主题相似度"分析模式。

2 研究思路与方法

在多位态的科学主题演化路径分析中,涉及到研究主题发展成熟程度预判、时间区间划分、研究主题识别和主题类型划分几个方面。

2.1 研究主题发展成熟度预判

科学技术的迅速发展引起科学知识量的加速度增加,科学文献作为科学知识的载体,它的变化可以作为直接反映科学发展的重要标志[20]。对科学信息增长规律进行研究能够揭示科学发展的特点和规律、预测文献量的增长趋势、判定科学发展阶段。著名科学史和科学学家库恩在《科学革命的结构》一书中提出科学发展模式是:常规科学在规范内的渐进积累－反常现象导致危机－新规范出现并引起科学革命－新的常规科学积累发展[21],是一个动态的历史过程和新陈代谢的革命过程,该过程呈现出"阶段革命"的特点。因此为判断科学主题发展阶段,笔者选择科学信息增长模型中阶段特征明显、影响最大的文献逻辑增长模型[22],即

$$F(t)=\frac{k}{1+ae^{-bt}}\quad(b>0)$$

其中:F(t)为t年文献累积量;k为t→∞时文献累积量;a和b为参数,前者用来确定成长曲线的位置,后者用来确定成长曲线的形状。

根据科学文献增长规律可将文献逻辑增长模型曲线分为4个阶段,分别为兴起、发展、成熟和衰退,如图1所示:

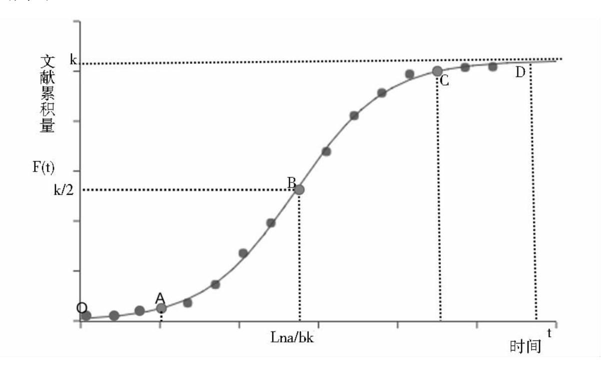


图1 文献逻辑增长模型曲线示例

注:横坐标Lna/bk为文献增长率最大临界点,经过此点文献增长率将逐渐减小

兴起(OA段):少量科学家针对某一方面进行原始基础理论研究,此时由于科学家基数较少且原始理论研究复杂,科学知识增长缓慢,作为知识载体的文献量保持缓慢增长。

发展(AB段):由于少数科学家的努力,基础理论得以突破,初步形成研究范式,吸引大量科学家进入,科学发展模式由"原始科学"转变为"常规科学",科学知识量以加速度方式增加,科学文献发表量呈指数形式增加。

成熟(BC段):随着研究的不断深入,大量问题得以解决,研究变得愈发困难,此时科学知识增速放缓,

常规学科发展迎来瓶颈。

衰退(CD 段):常规学科发展进入衰退期,科学问题几乎全部得以解决或研究遇到目前难以逾越的障碍,科学家纷纷退出该领域。在此阶段仍有少量科学家继续在该领域研究,仍有少量文献在此阶段产生。当这些科学家突破研究瓶颈或发现新的基础原理时,科学研究将再次进入由"原始科学"向"常规科学"的转变。在科学革命的推动下整个科学发展过程将按照上述规则周而复始地进行。

笔者利用文献逻辑增长模型对研究对象发展阶段进行预测,并将其扩展到作者和关键词的增长预测,以从学科规模(科学家累积量)和学科内容(关键词累积量)角度对研究领域发展阶段和发展趋势进行综合判定。

2.2 时间区间划分

进行主题演化路径研究时,需要重点考虑文献时间属性,根据样本数据特点选用恰当的方法对文献进行时间分区尤为重要。知名软件 CiteSpace[23] 和 Vos-viewer[24] 等均从时间视角展示主题演化情况,但上述软件以单个词为演化单元,无法展示主题之间的复杂演化关系。目前,常用的时间区间划分方法是 Time Line 方法[25] 和固定时间窗口方法[26]。由于 Time Line 方法复杂度较高,划分效果难以保证,笔者综合考虑文献量和文献发表时间用固定时间窗口方法对原数据进行时间分区。

2.3 研究主题识别

研究主题识别以每个时区关键词共现矩阵为基础,常用聚类算法包括谱聚类[27]、层次聚类[28]、Fast Newman[29] 和 Louvain[30] 算法等。笔者选用 Louvain 算法,该算法运行速度较快,适用于庞大网络的社群发现,同时采用启发式方式,能够克服传统 Modularity 类算法的局限,算法核心是最大化模块度 Q 值。Louvain 算法的基本原理是:首先将所有网络节点视为独立的社区,遍历每个节点的相邻节点,依据 Q 值将其归入相邻节点所在社区,直至每个节点所属社区不再发生变化;然后将社区视为独立节点,边的权重为两个节点内所有原始节点的边权重之和,重复上述步骤直至迭代到完全收敛。模块度 Q 值的计算如公式(1)所示:

$$Q = \frac{1}{2m}\sum_{i,j}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j)$$

$$\delta(u,v) = \{\frac{1 when: u == v}{0 else} \qquad \text{公式(1)}$$

其中,m 为网络中边的总数量,k_i 表示所有指向节点 i 的边的权重之和,k_j 同理,A_{ij} 表示节点 i 和 j 之间的边的权重,c_i 和 c_j 表示节点 i 和节点 j 所属的社区,如果 c_i,c_j 相同则为 $\delta(c_i, c_j) = 1$ 否则为 0。

2.4 主题类型划分

经过 Louvain 算法识别的研究主题尚不能充分体现这些研究主题所位于的发展阶段,笔者利用 J. Law 等[31] 人提出的战略坐标图(用来表示某个领域的主题类团现状的一种方法)来对研究主题的发展阶段进行区分。

实际上科学主题位于不同的发展阶段,在主题网络中也位于不同的位置。目标领域各主题并非完全处于同等地位,存在着核心－边缘、成熟－非成熟、增长－衰退－消失等不同类型主题。因此,笔者分别用向心度(T_c)和密度(T_d)两个指标(见公式(2)和公式(3))[32] 测度主题聚类之间联系的强弱和主题聚类内部联系的强弱。T_c 值越大,意味着主题聚类之间的联系越紧密,表明该主题受到广泛关注,易成为研究热点,该主题聚类在所属的研究领域中发挥核心作用。T_d 值越大,意味着主题聚类内部的联系越紧密,表明该主题相对成熟。分别以 T_c 和 T_d 为横纵坐标轴,T_c 和 T_d 的均值为坐标原点构建科学主题发展程度的四象限图谱(见图 2),每个象限表示一种主题类型,象限内的不同圆圈表示不同主题,圆圈之间的实线表示目标主题与其他主题之间联系紧密,即目标主题在整个研究工作中处于核心地位;虚线表示目标主题与其他主题之间联系松散,即目标主题在整个研究工作中处于边缘地位。目标主题圆圈内部各个节点表示该主题所包含的关键词,关键词之间的实线表示该主题内部联系紧密,属于成熟型主题,虚线表示该主题内部联系松散,属于非成熟型主题。因此,位于第一象限的主题向心度和密度均较高,属于关联广泛的成熟主题;位于第二象限的主题向心度较低、密度较高,属于相对独立的成熟主题;位于第三象限的主题向心度和密度均较低,是整个领域的边缘主题,内部结构比较松散,研究尚不成熟。因此可能是衰退后的主题,也可能是新兴的主题,目前尚处于发展初期,主题内部和主题之间的联系较为松散,至于属于哪种类型还需要结合实际进行专业判断;位于第四象限的主题向心度较高、密度较低,属于有着广泛联系的不成熟主题,具有较大的发展潜力。

$$T_c = \frac{\sum_{i\in\varphi_s, j\in(\varphi-\varphi_s)} E_{ij}}{N-n} \qquad \text{公式(2)}$$

$$T_d = \frac{\sum_{i,j\in\varphi_s} E_{ij}}{n-1}, \ (i \neq j) \qquad \text{公式(3)}$$

其中,E_{ij}是关键词的共现频次,n 是某一聚类主题中关键词的数目,N 是共词矩阵中所有关键词的数目,φ_s 指的是某一个主题,φ 指的是全部主题。

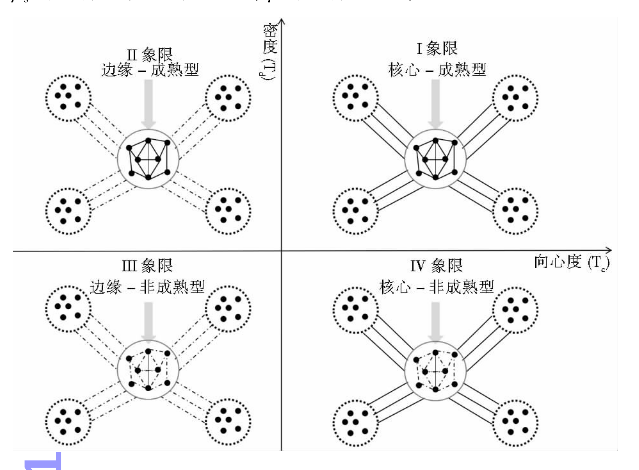


图 2 战略坐标图的四个象限解析

2.5 主题演化路径

战略坐标图对主题类型划分具有优势,但该图仍属于静态图,无法体现不同时间段各主题之间的联系,无法研究相同或不同类型主题之间的演化、转化等趋势。基于此,笔者提出通过计算相邻时区战略坐标图中主题间的相似度,用以表征主题演化方向。相似度计算采用余弦相似度,即将每个主题包含的关键词映射为向量,通过计算相邻窗口不同主题之间的向量夹角余弦值来判定两个主题的相似性,该值越大表明相似度越高,见公式 4(其中 A 和 B 为主题对应的向量)。同时结合战略坐标自身特性,根据不同时区主题在战略坐标图中所处象限位置的变化揭示主题发展类型演化趋势,如稳定型主题、新兴型主题、成熟型主题或衰退型主题等。

$$\cos(\theta)=\frac{A\cdot B}{\|A\|\ \|B\|}=\frac{\sum_{i=1}^{n}A_i\times B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2}\times\sqrt{\sum_{i=1}^{n}(B_i)^2}}$$

公式(4)

3 3D 打印技术的主题演变

本文以 3D 打印技术领域为案例研究对象,数据来源为 Web of Science 数据库核心集。综合考虑查全率与查准率,制定如下检索策略:TI = (“Additive Manufact * ” OR “Rapid Prototyp * ” OR “3D Print * ” OR “three dimensional * print * ” OR “Solid Free * form Fabricat * ”),语种限定为 English,文献类型限定为 Article OR Review,时间跨度为所有年份(检索时间 2020 年 6 月 24 日),共检出文献 15 233 篇,其中含关键词字段文献数 12 028 篇,以此作为数据源。通过上述方法进行多位态研究主题识别及其演化路径分析,研究思路见图 3,具体论述如下。

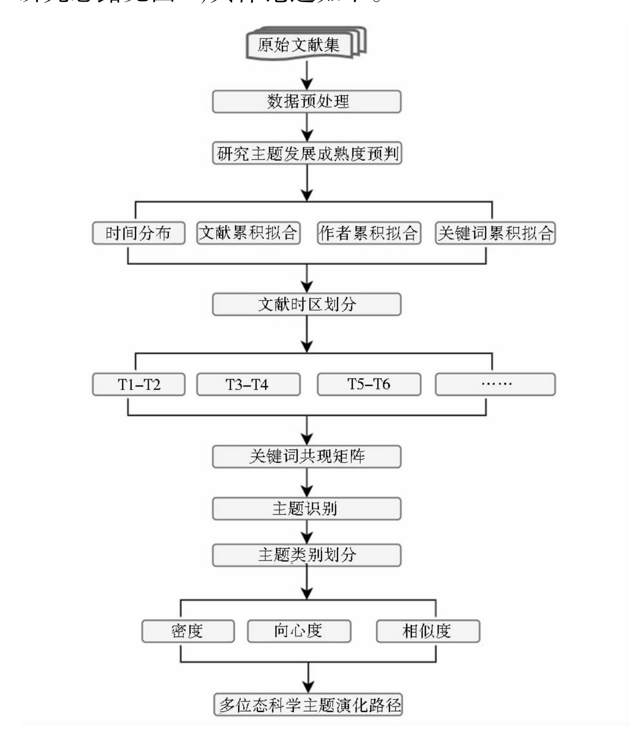


图 3 多位态科学主题识别及其演化路径研究思路

3.1 3D 打印技术的发展阶段

3D 打印的相关论文数量、论文关键词数量和作者数量都呈现快速增长的态势,尤其是 2012 年之后表现更为明显(见图 4a),依据这 3 个指标都与逻辑增长模型曲线有着较好的拟合,拟合优度 R2 分别为 0.948、0.957、0.960(见图 4b),能够有效地预测 3D 打印技术的发展阶段。结合文献逻辑增长模型曲线四阶段划分方法可知 3D 打印技术正处于第二个发展阶段(AB 段),距离第二阶段结束仅有 2 年时间。在此期间关于 3D 打印的科学文献仍会保持指数增长趋势,大量科研人员将继续涌入该领域,研究内容也将不断扩展。因此,我国应把握这两年的机遇期,力争在 3D 打印领域占领制高点。

3.2 3D 打印技术的研究主题识别

首先采用固定时间窗口方法对 3D 打印文献进行时间区间划分,考虑到 3D 打印技术发展初期较长时间内文献量较少,故将 1991 – 2009 年划分为第一时间区间、2010 – 2014 年为第二时间区间、2015 – 2016 年为第三时间区间、2017 – 2018 年为第四时间区间、2019 – 2020 年为第五时间区间,各时间区间具体文献数量见表 1。

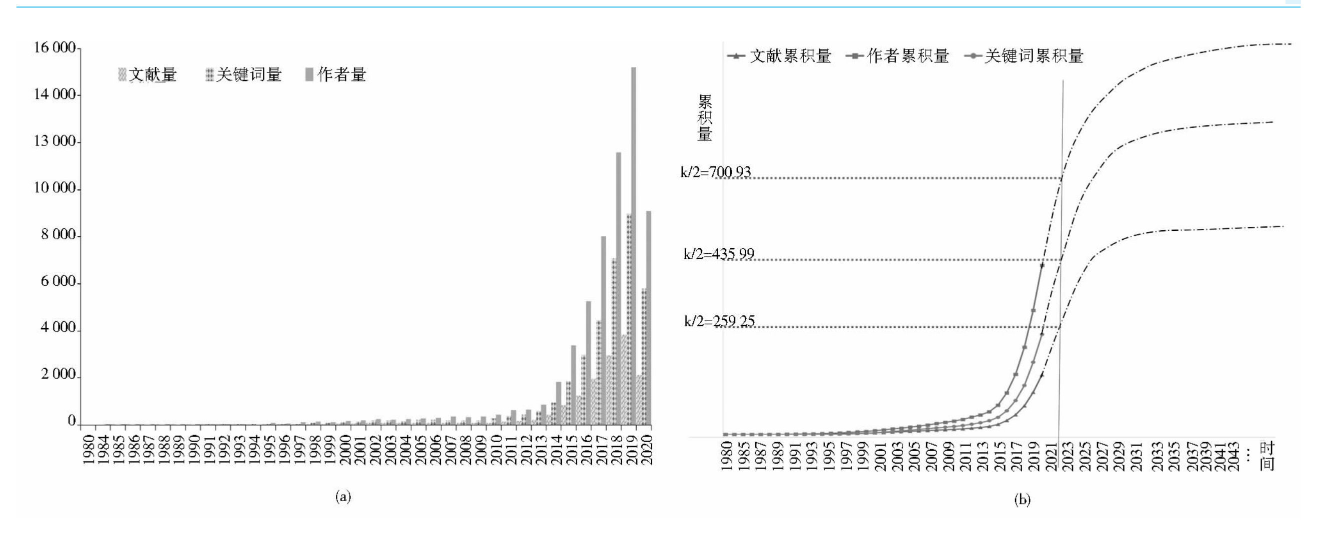


图 4　3D 打印技术发展阶段预测

表 1　时间窗口主题网络指标统计

时间窗口	文献量/篇	网络节点数	边数	平均度	密度	平均聚类系数	主题(成员>=3)	研究主题
1991－2009	692	35	39	4.871 8	0.128 2	0.315 4	5	3D 打印工艺(SLA/SLS/FDM)、计算机辅助设计、组织工程支架、逆向工程、机械性能
2010－2014	718	41	109	5.317 1	0.132 9	0.371 4	3	3D 打印工艺与计算机辅助设计、组织工程支架、金属 3D 打印
2015－2016	1 505	107	390	7.289 7	0.068 8	0.360 1	6	生物 3D 打印、3D 打印工艺、金属 3D 打印、计算机辅助设计、微流控、3D 打印材料
2017－2018	3 993	376	2 391	12.718 1	0.033 9	0.340 6	8	塑料 3D 打印、金属 3D 打印与激光 3D 打印、生物 3D 打印、3D 打印材料、计算机辅助设计、粘结剂喷射 3D 打印、电化学、3D 混凝土打印
2019－2020	5 120	538	4 067	15.119 0	0.028 2	0.328 6	7	熔融沉积建模技术与聚乳酸、金属、激光、电弧 3D 打印、生物 3D 打印、光固化立体造型和选择性激光烧结工艺、计算机辅助设计、3D 打印可持续发展、3D 打印模型(医学)

3.2.1　关键词共现矩阵构建

提取每个时间窗口高频关键词共现对(前 20 对关键词对,去除含检索词对),如图 5。其中,子图横轴数字与纵轴关键词一一对应,共现频次为对角线数字。随着时间发展,各区间的空白区逐渐减少,说明前 20 对关键词对之间的联系愈发紧密。在 1991－2009 年、2010－2014 年两个时间段中,支架与组织工程共现频次远远高于其他关键词对,表明早期的研究热点之一是利用 3D 打印技术在骨、软骨、血管等组织工程领域构建符合组织器官结构特点的支架。在 2015－2016 年时间段中,支架与组织工程共现频次仍然最高,此外该阶段机械性能与微结构和组织工程与生物打印开始频繁共现,表明 3D 打印研究热点随着时间的发展研究领域不断扩展。在 2017－2018 年时间段中,关于机械性能与微结构的共现频次已上升为第一,其次是支架与组织工程、激光 3D 打印与微结构、组织工程与聚乳酸材料、组织工程与机械性能、微机构与钛合金等,3D 打印热点进一步扩展到打印工艺(激光 3D 打印)、打印材料(聚乳酸、钛合金)。在 2019－2020 年时间段中,机械性能与微结构共现频次仍保持领先,其次是熔融沉积建模技术与聚乳酸、电弧增材制造与微结构、支架与组织工程、机械性能与电弧增材制造、激光 3D 打印与微结构、熔融沉积建模技术与机械性能。表明近期关于 3D 打印工艺的研究仍然占据主导地位,其中熔融沉积建模 3D 打印、电弧 3D 打印和激光 3D 打印尤为突出。

3.2.2　3D 打印主题提取

学术论文的关键词是论文内容的提示符,也是作者学术思想与观点的高度凝练。两个或者多个关键词在同一篇文章中的同时出现称为关键词共现。关键词共现通过描述关键词与关键词之间的关联与结合,可以揭示学术领域研究内容的内在相关性和学科领域的微观结构,即可以用来揭示某个研究领域的研究热点。笔者选取频次大于等于 5 的关键词构建共现矩阵,并在此基础上利用 Python 语言 community 包中的 Louvain 算法进行主题聚类并计算网络基本指标(见表 1),网

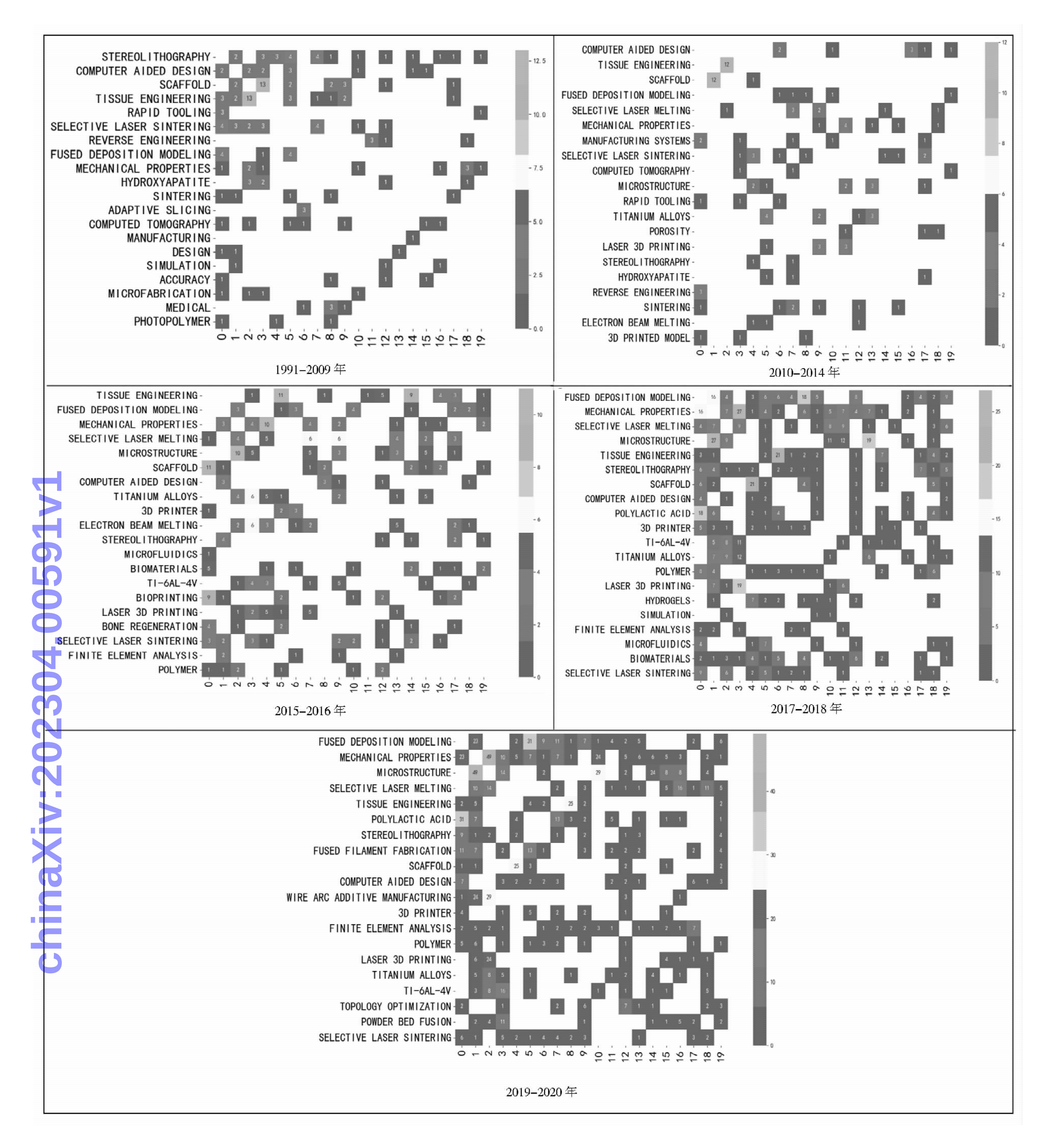


图 5　3D 打印各时间区间关键词共现

络节点数、边数和平均度逐渐增加，密度和平均聚类系数呈下降趋势，网络整体连接较松散，关键词之间仍有较大的关联空间。基于时间窗口的关键词聚类可清晰展示 3D 打印领域研究主题的变迁：3D 打印工艺和计算机辅助设计贯穿 3D 打印研究始终；塑料、金属和生物材料是 3D 打印的主要材料来源；在生物应用方面，3D 打印由在组织工程支架方面的应用逐渐扩展到生物领域方方面面；近年来关于电弧 3D 打印和 3D 打印废品回收与可持续发展成为新兴主题。

3.3　主题类型划分与演化路径

上述聚类可有效识别各个时间区间的主题，但其将所有主题同等对待，无法有效区分各主题的地位，同时亦无法揭示相邻主题间的关系。基于此，笔者在聚类基础上通过密度和向心度两个指标对主题类别进行划分，通过余弦相似度指标（阈值 > = 0.1）对主题间具体演化关系进行度量（见图 6）。在图 6 中，不同象限的主题具有不同的地位，根据战略坐标划分方法可得第一象限代表核心成熟型主题、第二象限代表边缘

成熟型、第三象限代表边缘非成熟型、第四象限代表核心非成熟型;不同时间区间主题间的联系代表主题演化关系,线条上的数字表示两主题相似度大小,值越大表示继承前一主题的内容越多。

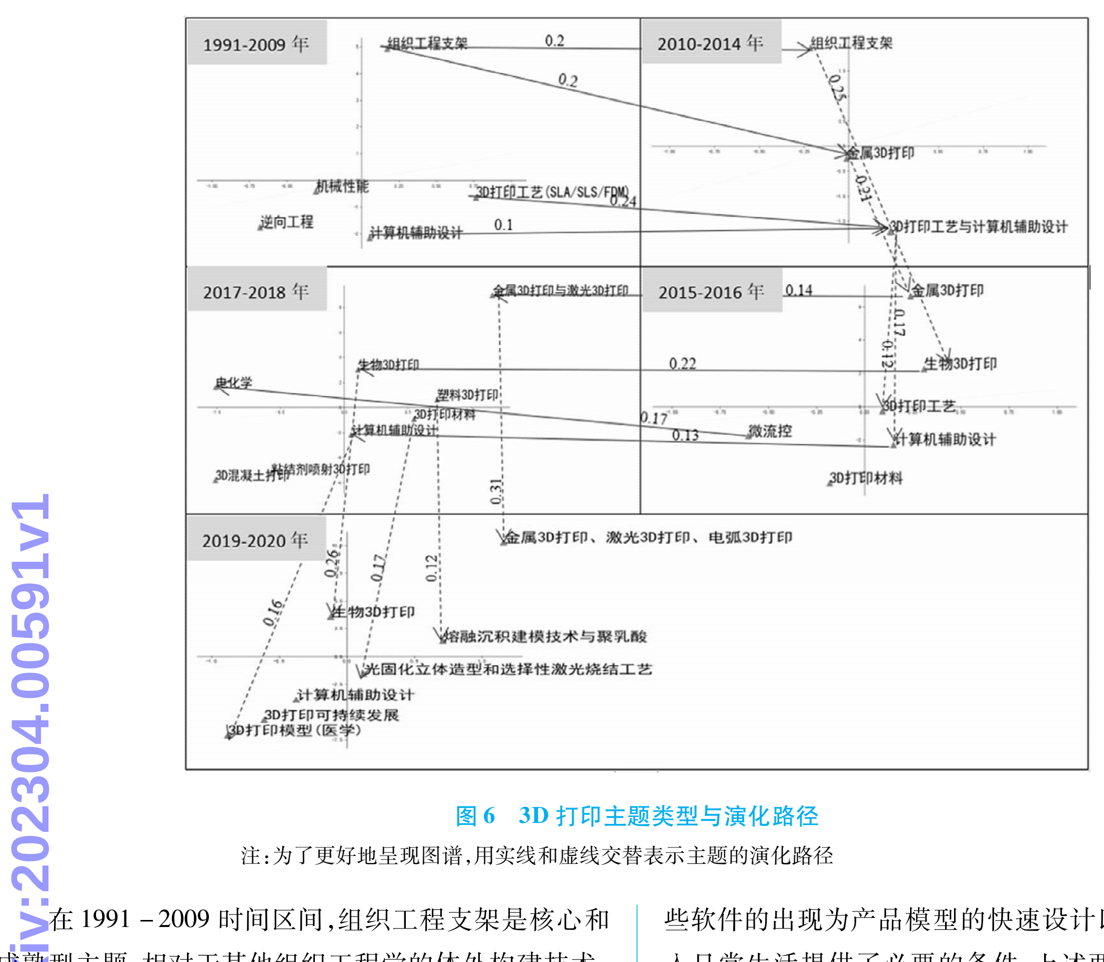


图 6　3D 打印主题类型与演化路径

注:为了更好地呈现图谱,用实线和虚线交替表示主题的演化路径

在 1991 – 2009 时间区间,组织工程支架是核心和成熟型主题,相对于其他组织工程学的体外构建技术,3D 打印技术特别适合复杂结构支架的成型,在制造具有多孔和复杂微观结构的复合植入物方面具有独特优势[33]。组织工程支架主题在下一时间区间进行了分化,一方面继续沿着自身主题发展,另一方面向金属 3D 打印方向演进;新兴或衰退型主题是机械性能和逆向工程,上述两个主题在下一阶段并未出现;发展型主题是 3D 打印工艺和计算机辅助设计,该阶段打印工艺主要包括利用紫外光逐层照射光敏树脂液面以固化成型的光固化立体造型技术(Stereo Lithography Apparatus, SLA)、激光将粉末颗粒局部融合在一起的选择性激光烧结技术(Selective Laser Sintering, SLS)和将融化挤出的材料逐层地在工作台上堆积成型的熔融沉积建模技术(Fused Deposition Modelling, FDM);计算机辅助设计是通过计算机为 3D 打印机生成具有特定格式的 3 维模型数据用于打印成型产品,其是产品模型设计不可或缺的支撑技术,常用的建模软件包括 Solid works、Alias、Rhino、Solid works、Sketch up 和 C4D 等,这些软件的出现为产品模型的快速设计以及 3D 打印进入日常生活提供了必要的条件,上述两个主题在下一阶段融合成为同一主题。

在 2010 – 2014 时间区间,该时间段内主题较少,没有核心成熟型主题,组织工程支架由前期的第一象限变为目前的第二象限,即由核心主题变为边缘主题,但其在下一时间区间演化为生物 3D 打印,表明该主题研究范围得到了扩展,从对组织工程支架的打印扩展到生物领域的其他诸多方面;金属 3D 打印处于第三象限,属于新兴型主题,其在下一时间区间继续沿着金属 3D 打印路径发展,金属 3D 打印常用的材料有钛合金、镁铝合金、镍基合金、铜基合金、不锈钢、钴铬合金等,1995 年弗劳恩霍夫研究所发明的选择性激光熔化技术(Selective Laser Melting, SLM)使用金属粉末代替 SLS 中的高分子聚合物作为粘合剂,打印出的成型产品精度和力学性能优于 SLS 工艺;3D 打印工艺与计算机辅助设计处于第四象限,与前期相比该主题所处象限位置并未发展改变,属于发展型主题,其在下一时间区间分化为 3D 打印工艺、计算机辅助设计两个主题。

在2015－2016时间区间，金属3D打印由前期第三象限变为第一象限，即由新兴主题快速发展为核心－成熟型主题；生物3D打印由处于第二象限的边缘－成熟型主题组织工程支架发展为第一象限的核心－成熟型主题生物3D打印；3D打印工艺与计算机辅助设计仍处于第四象限，与前期相比该主题所处象限位置并未发展改变，属于发展型主题；新出现了3D打印材料和微流控两个主题，其中打印材料以聚乳酸（PLA）为主要研究对象，PLA是可降解材料，一般是作为线状材料使用，3D打印成型利用其融化和冷却成型的特点，即加热到一定温度时会融化，温度降低又会重新凝固；在微流控主题方面，随着3D打印精度的不断提升，利用高精度3D打印机制造微流控芯片越来越可行与方便，该技术能够显著简化微流控芯片的加工过程，同时在打印材料的选择上也较为灵活。

在2017－2018时间区间，金属3D打印和生物3D打印继续处于第一象限，属于核心－成熟型主题，是前期同一主题的继续延伸，该阶段的核心－成熟型主题增加了塑料3D打印；3D打印材料由前期的第三象限新兴型主题变为第四象限发展型主题，打印材料进一步得到了扩展，包括聚合物、生物材料、复合材料、陶瓷、碳纳米管、氧化锆、智能材料、多孔材料、热塑、硅酮、金属、软材料、聚偏氟乙烯、热塑性聚合物、碳纤维、丙烯酸酯、聚二甲基硅氧烷、树脂等。由于3D打印材料主题扩展速度较快，新关键词增加较多，致使与前期同一主题相似度偏低；计算机辅助设计仍然处于第四象限，自1991年以来主题类型尚未发生改变，主题发展稳定性较强；电化学主题由前期微流控主题演化形成，由第三象限变为第二象限，主题成熟方面有所增加，但仍属于边缘主题；3D混凝土打印处于第三象限，属于新兴型主题，3D打印在建筑领域应用已初具规模，如2017年6月以混凝土3D打印著称的荷兰3D打印公司CyBe Construction利用3D打印技术为迪拜打印出了全球首个R&Drone实验室，在节省时间和材料方面表现出强大的潜力；粘结剂喷射3D打印技术也是该阶段的新兴主题，GE Aviation、Desktop Metal、惠普、Markforged、3DEO等公司在该领域均有较强的实力，该技术通过材料喷射和烧结工艺的相互结合来生产完全密度的金属零部件，速度更快，如GE Additive的粘结剂喷射技术于2017年首次亮相，与更常见的基于激光的3D打印技术相比，粘结剂喷射可以使打印速度快60到100倍。由于用于粘结剂喷射技术的材料一般不贵且可100%回收未使用的粉末，生产成本相对更低。

在2019－2020时间区间，该阶段主题划分更为细致，如3D打印工艺单独成为研究主题，熔融沉积建模和聚乳酸由前期塑料3D打印演化而来，仍处于第一象限，属于核心－成熟型主题；光固化立体造型和选择性激光烧结技术由3D打印材料演化而来，仍处于第四象限，属于核心－不成熟型主题，有继续发展的潜力；金属3D打印、激光3D打印和电弧3D打印由前期相同主题演化而来，处于第一象限，属于核心－成熟型主题，且其向心度和密度远远大于其他主题，说明目前的研究热点主要集中于金属、激光和电弧3D打印；3D生物打印由第一象限变为第二象限，由核心－成熟型变为边缘－成熟型，此方面研究有衰退的趋势；3D打印模型（医学）由计算及辅助设计演化而来，该技术为精准医疗提供了条件，通过打印患者器官、骨等三维模型，可以预先模拟处理方法以选择最佳的治疗手术方法，对提高手术准确度具有重要意义，同时利用打印的模型可有效对年轻医生进行培训。此外，3D打印模型还可用于假肢、矫正夹具等各类可穿戴设备；计算及辅助设计在此时间段处于第四象限，结合前期主题发展情况，可知该主题属于衰退型主题；3D打印可持续发展处于第三象限，属于新兴型主题。3D打印是通过增材的方式以减少资源的浪费，属于环保型技术，但与此同时3D打印的材料以及打印过程产生的部分废料、废气如果处理不当也存在着一定的安全隐患。如美国加州大学河滨分校的研究人员发现，一些商业3D打印机制造的物品对某些鱼类胚胎来说具有致命毒性。因此，如何鉴别有毒材料以及如何处理废料实现可持续发展值得进一步关注。

4 结语

主题演化分析是研究学科发展的一种重要方法，已得到研究人员的广泛关注，但目前的研究将不同主题同等对待，无法有效区别不同主题的重要程度。基于此，笔者提出的多位态科学主题识别及其演化路径方法能够弥补这一缺陷，便于详细展示领域发展过程中核心、边缘、成熟、非成熟等类型主题的出现、发展、衰退趋势以及不同类型主题的动态交互过程，对把握学科领域的发展过程和未来方向，尤其是对颠覆性技术的识别与发展战略的制定具有重要的意义。

参考文献：

[1] TONY H, STEWART T, KRISTIN T, et al. 第四范式：数据密集型科学发现[M]. 潘教峰，张晓林，安培浚，等译. 北京：科学出版社，2012.
[2] 敦帅,陈强. 中国企业可持续发展研究：态势演进与主题演化[J]. 科学管理研究,2020,38(1):94－100.
[3] 杜丽君. 学科交叉视角下的信息检索研究主题演化分析——以情报学和计算机科学为例[J]. 信息技术与信息化,2020(1):178－183.
[4] 王璐瑶,曲冠楠,李楠,等. 中国科技管理研究的贡献、演进与展望——《科学学与科学技术管理》创刊 40 年文献计量与知识图谱分析[J]. 科学学与科学技术管理,2020,41(5):3－21.
[5] WU Y, JIN X, XUE Y Z, et al. Evaluation of research topic evolution in psychiatry using co-word analysis[J]. Medicine, 2017, 96(25): 7349.
[6] ONYANCHA O B. Forty-five years of lis research evolution, 1971－2015: an informetrics study of the author-supplied keywords[J]. Publishing research quarterly, 2018, 34(3): 456－470.
[7] HAN X Y. Evolution of research topics in lis between 1996 and 2019: an analysis based on latent dirichlet allocation topic model[J]. Scientometrics, 2020, 125(3): 2561－2595.
[8] HUANG C, YANG C, WANG S T, et al. Evolution of topics in education research: a systematic review using bibliometric analysis[J]. Educational review, 2020, 72(3): 281－297.
[9] 牛力,杜丽华. 国内档案学学科主题网络构建与演化分析研究[J]. 档案学研究,2020(2):12－18.
[10] 王晓光,程齐凯. 基于 NEViewer 的学科主题演化可视化分析[J]. 情报学报,2013,32(9):900－911.
[11] 谭春辉,熊梦媛. 基于 LDA 模型的国内外数据挖掘研究热点主题演化对比分析[J]. 情报科学,2021,39(4):174－185.
[12] 颜端武,苏琼,张馨月. 基于时序主题关联演化的科学领域前沿探测研究[J]. 情报理论与实践,2019,42(7):144－150.
[13] 周耀林,柴昊,赵跃. 国际图情领域大数据研究现状与趋势探析[J]. 图书馆杂志,2019,38(12):16－27,44.
[14] TANCOIGNE E, BARBIER M, COINTET J P, et al. The place of agricultural sciences in the literature on ecosystem services[J]. Ecosystem services, 2014, 10(SI): 35－48.
[15] 赵跃,乔健. 改革开放 40 年来中国档案学研究的全景透视——基于研究主题的挖掘与演化分析[J]. 档案学研究,2019(3):44－54.
[16] RULE A, COINTET J P, BEARMAN P S, et al. Lexical shifts, substantive changes, and continuity in state of the union discourse, 1790－2014[J]. Proceedings of the National Academy of Sciences of the United States of America, 2015, 112(35): 10837－10844.
[17] MATOS F L , ROSS S W , HUVENNE V A I , et al. Canyons pride and prejudice: exploring the submarine canyon research landscape, a history of geographic and thematic bias[J]. Progress in oceanography, 2018, 169(12):6－19.
[18] MARVUGLIA A, HAVINGA L, HEIDRICH O, et al. Advances and challenges in assessing urban sustainability: an advanced bibliometric review[J]. Renewable & sustainable energy reviews, 2020, 124(3): 1－12.
[19] 岳丽欣,周晓英,陈旖旎. 期刊论文核心研究主题识别及其演化路径可视化方法研究——以我国医疗健康信息领域期刊论文为例[J]. 图书情报工作,2020,64(5):89－99.
[20] 王伟. 信息计量学及其医学应用[M]. 北京:人民卫生出版社,2014.
[21] 王海山. 从波普尔、库恩的科学观看科学发展的规律性[J]. 科学学研究,1984(1):34－40.
[22] 何荣利. 关于科学文献增长模型的思考[J]. 情报杂志,1994(6):40－41,64.
[23] CHEN C. Citespace II: detecting and visualizing emerging trends and transient pat terns in scientific literature[J]. Journal of the American Society for Information Science and Technology, 2006, 57(3): 359－377.
[24] VAN ECK N J , WALTMAN L. Software survey: vosviewer, a computer programfor bibliometric mapping[J]. Scientometrics, 2010, 84(2) : 523－538.
[25] 吴斌, 王柏, 杨胜琦. 基于事件的社会网络演化分析框架[J]. 软件学报, 2011, 22(7): 1488－1502.
[26] 钱铁云, 李青, 许承瑜. 面向科技主题发展分段的社区核心圈技术[J]. 计算机科学与探索, 2010, 4(2): 170－179.
[27] ULRIKE V L. A tutorial on spectral clustering[J]. Statistics and computing, 2007,17 (4): 1－32.
[28] HAN J W, KAMBER M. Data mining concepts and techniques [M]. New York: Academic Press, 2001.
[29] NEWMAN M E J. Fast algorithm for detecting community structure in networks[J]. Physical review e, 2004, 69(6): 1－15.
[30] BLONDEL V D, GUILLAUME J L, LAMBIOTTE R, et al. Fast unfolding of communities in large networks[J]. Journal of statistical mechanics: theory and experiment, 2008,9(10): 1－12.
[31] LAW J, BAUIN S , COURTIAL J P , et al. Policy and the mapping of scientific change: a co-word analysis of research into environmental acidification[J]. Scientometrics, 1988, 14(3/4):251－264.
[32] LEE B, JEONG Y I. Mapping korea's national R&D domain of robot technology by using the co-word analysis[J]. Scientometrics, 2008, 77 (1) :3－19.
[33] 田冶,曾庆慧,胡相华,等. 3D 打印技术及在组织工程领域的研究进展[J]. 中国医疗器械信息,2015,21(8):7－12.

作者贡献说明：

王康：论文构思、数据处理、论文撰写、论文修改；
陈悦：论文构思、论文修改；
高继平：数据处理与分析；
潘云涛：研究方法与结果分析。

Research on the Method of Multi-position Research Themes Recognition and Evolution Path

Wang Kang[1] Gao Jiping[2] Pan Yuntao[2] Chen Yue[1]
[1] Institute of Science of Science and S&T Management & WISE Lab, Dalian University of Technology, Dalian 116024
[2] Institute of Scientific and Technical Information of China, Beijing 100038

Abstract: [Purpose/significance] The evolution path of scientific themes is of great significance for understanding the process of scientific development and predicting future development trends. In view of the defect that the existing research treats the topics on the evolution path equally, a new method of multi-position scientific topic identification and its evolution path is proposed. [Method/process] This method divides the topics of each time interval into four types: core-mature, edge-mature, edge-immature, and core-immature based on centripetal degree and density, and uses cosine similarity to divide different time interval themes are related to show the dynamic cross-evolution relationship between different types of scientific themes. [Result/conclusion] Taking the literature of 3D printing as an example, the development process of 3D printing technology was measured from four aspects: technology development stage, theme recognition, theme type division and theme evolution path. The results proved that this method is effective for scientific theme recognition and its evolution path display.

Keywords: topic recognition topic evolution strategic coordinates topic similarity 3D printing